

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

FAKULTA ELEKTROTECHNIKY
A KOMUNIKAČNÍCH TECHNOLOGIÍ

BAKALÁŘSKÁ PRÁCE



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA ELEKTROTECHNIKY

A KOMUNIKAČNÍCH TECHNOLOGIÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

ÚSTAV AUTOMATIZACE A MĚŘICÍ TECHNIKY

DEPARTMENT OF CONTROL AND INSTRUMENTATION

IDENTIFIKACE OSOBY

PERSON IDENTIFICATION

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

AUTOR PRÁCE

AUTHOR

Stanislav Svědřoh

VEDOUcí PRÁCE

SUPERVISOR

Ing. Miloslav Richter, Ph.D.

BRNO 2020

Bakalářská práce

bakalářský studijní program **Automatizační a měřicí technika**

Ústav automatizace a měřicí techniky

Student: Stanislav Svědřih

ID: 203424

Ročník: 3

Akademický rok: 2019/20

NÁZEV TÉMATU:

Identifikace osoby

POKYNY PRO VYPRACOVÁNÍ:

- 1) Nastudujte metody 3D rekonstrukce a popisu obličeje.
- 2) Nastudujte možnosti využití 3D a 2D dat pořízených pomocí senzoru Kinect k identifikaci osoby podle obličeje.
- 3) Nastudujte možnosti senzoru Kinect k získání dat a realizujte algoritmy pro jeho ovládání.
- 4) Navrhněte a realizujte algoritmy pro detekci a rozpoznání obličeje a identifikaci osoby pomocí dat ze senzoru Kinect. Stanovte požadované vlastnosti vstupních dat a formát dat výstupních.
- 5) Navrhněte a realizujte měřicí pracoviště a proveďte pořízení datových sad pro ověřování navržených algoritmů.
- 6) Zhodnoťte možnosti a kvalitu navrženého řešení k identifikaci osoby.

DOPORUČENÁ LITERATURA:

Gonzalez R.C., Woods R.E.: Digital Image Processing, 4th edition, Pearson, 2017, ISBN 978-0133356724

Termín zadání: 3.2.2020

Termín odevzdání: 8.6.2020

Vedoucí práce: Ing. Miloslav Richter, Ph.D.

doc. Ing. Václav Jirsík, CSc.
předseda rady studijního programu

UPOZORNĚNÍ:

Autor bakalářské práce nesmí při vytváření bakalářské práce porušit autorská práva třetích osob, zejména nesmí zasahovat nedovoleným způsobem do cizích autorských práv osobnostních a musí si být plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č.40/2009 Sb.

ABSTRAKT

Práce se zabývá problematikou identifikace osob na základě obličeje. Cílem je vytvořit kompletní řešení identifikace pro účely autentizace za pomoci levného senzoru Microsoft Kinect v2. Práce obsahuje souhrn nejpoužívanějších algoritmů k popisu dat obsahujících lidský obličej a jejich porovnání. Další částí je PC program obsluhující zmiňovaný senzor pro sběr obličejových dat, jejich ukládání, extrakce popisných bodů a jejich porovnávání. Výstupem programu je procentuální shoda skenovaného obličeje s obličejem v databázi

KLÍČOVÁ SLOVA

Identifikace obličeje, Microsoft Kinect v2, autentizace, program, C#, algoritmy, 3D model, MATLAB

ABSTRACT

This thesis focuses on the person identification based on faces. The aim is to create a working solution of person identification for the authentication purposes utilizing the Microsoft Kinect v2 sensor. It contains a summary of the most used algorithms for facial data description and their comparison. Thesis also proposes a PC program that controls the sensor and collects the facial data, saves them, manages keypoint detection and their comparison.

KEYWORDS

Facial identification, Microsoft Kinect v2, authentication, program, C#, algorithms, 3D model, MATLAB

Bibliografická citace

SVĚDIROH, Stanislav. *Identifikace osoby* [online]. Brno, 2020 [cit. 2020-06-07]. Dostupné z: <https://www.vutbr.cz/studenti/zav-prace/detail/126973>. Bakalářská práce. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav automatizace a měřicí techniky. Vedoucí práce Miloslav Richter.

Prohlášení

„Prohlašuji, že svou závěrečnou práci na téma Identifikace osoby jsem vypracoval samostatně pod vedením vedoucího závěrečné práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autor uvedené závěrečné práce dále prohlašuji, že v souvislosti s vytvořením této závěrečné práce jsem neporušil autorská práva třetích osob, zejména jsem nezasáhl nedovoleným způsobem do cizích autorských práv osobnostních a jsem si plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

V Brně dne **8. června 2020**

..... podpis autora

Poděkování

Děkuji vedoucímu práce Ing. Miloslavu Richterovi, Ph.D. za cenné rady, které mi velmi pomohly při jejím vypracování. Dále bych rád poděkoval Janu Štanclovi, Jiřímu Pražákovi, Sylvě Profousové a Pavlu Duškovi za poskytnutí dat pro testování.

V Brně dne **8. června 2020**

..... podpis autora

Obsah

1	Úvod	1
2	Metody 3D rekonstrukce obličeje	3
3	Popis obličejových dat	5
3.1	Holistický přístup	5
3.1.1	ICP (Iterative closest point)	5
3.1.2	PCA (Principal component analysis)	6
3.2	Lokální přístup	7
3.2.1	Metody klíčových bodů (Keypoint-based methods)	8
3.2.2	Křivkové metody (Curve-based methods)	10
3.3	Hybridní přístup	11
4	Microsoft kinect V2	12
4.1	RGB Kamera	13
4.2	Infračervená kamera	13
4.3	Datové proudy	13
4.3.1	Hlubkový datový proud	14
4.4	Kinect Fusion	15
4.4.1	Snímání prostředí	16
4.4.2	Vyhodnocení polohy senzoru	16
4.4.3	Aktualizace mračna bodů	16
4.4.4	Výpočet výsledného modelu	16
5	Obslužný program	17
5.1	Počáteční inicializace senzoru	18
5.2	Přejímání dat z datových proudů	18
5.3	Zpracování přijatých dat	18
5.4	Uložení 3D rekonstrukce	19
5.5	Automatizace sběru dat	19
6	Extrakce popisných bodů	20
6.1	Předzpracování vstupních dat	20
6.1.1	Počáteční filtrace	21
6.1.2	Segmentace obličeje	21
6.2	Detekce nosu a normalizace dat	23
6.3	Extrakce obličejových křivek	23
6.4	Extrakce tvaru nosu	25
6.5	Extrakce tvaru očních koutků	26
7	Porovnání popisných bodů	27
7.1	Porovnání tvaru nosu a očních koutků	27
8	Fáze testování	29
8.1	Akvizice základní datové sady	29

8.2	Testování kvality popisných bodů a jejich extrakce	30
8.3	Stanovení vztahu pro výslednou procentuální shodu	30
8.4	Nutné podmínky pro správnou funkčnost algoritmu	31
9	Použitelnost systému v praxi	32
	Závěr	33
	Literatura	34
	Seznam příloh	36

Seznam obrázků

Obr. 1: 3D Rekonstrukce obličeje z fotografie algoritmem navrženou v [1]	3
Obr. 2: 3D rekonstrukce obličeje pomocí algoritmu Kinect Fusion [12]	4
Obr. 3: Vizualizace zarovnání 3D modelů pomocí ICP metody [3]	6
Obr. 4: Příklad zarovnání obličeje s referenčním obličejem pro klasifikaci metodou PCA [4]	7
Obr. 5: Detekované body síťovou metodou (vpravo) a rozložení jejich okolí pro účely popisu (vlevo) [13]	8
Obr. 6: Ukázka extrakce bodů s největší, zakřivením pomocí prahování [7]	9
Obr. 7: Vizualizace unikátního trojúhelníku pro popis obličeje [7]	9
Obr. 8: Vizualizace konturových metod hloubkovým způsobem (a), způsobem konstantní vzdálenosti (b) [2]	10
Obr. 9: vizualizace profilové metody (vpravo) s binární maskou (vlevo) [11]	11
Obr. 10: Rozložení snímačů v senzoru Microsoft Kinect V2 [14]	12
Obr. 11: Příklad inicializace senzoru Microsoft Kinect V2	13
Obr. 12: Metody přejímající data z jednotlivých datových toků asynchronně	14
Obr. 13: Příklad 3D rekonstrukce obličeje z jednoho hloubkového snímku (vlevo), z více snímků (vpravo)	15
Obr. 14: Blokové schéma procesu rekonstrukce scény pomocí algoritmu Kinect Fusion [12]	16
Obr. 15: Ukázka obslužného programu	17
Obr. 16: Vizualizace vstupních dat před filtrací (vlevo) a po filtraci (vpravo)	21
Obr. 17: Segmentace obličejových dat pomocí algoritmu kmeans	22
Obr. 18: Vizualizace obličejových dat před segmentací (vlevo) a po segmentaci (vpravo)	23
Obr. 19: Extrakce hřebene nosu (červená) a okolí špičky nosu (modrá)	25
Obr. 20: Detekce očních koutků z derivace obličejové křivky (vlevo) a jejich okolí (vpravo)	26
Obr. 21: Vizualizace porovnání okolí nosu pro výpočet míry shody	27
Obr. 22: Vizualizace porovnání tvaru očních koutků	28

1 ÚVOD

V současné době se technologický průmysl zaměřuje a bere velmi vážně otázku soukromí a ochrany dat. Veškerá data na internetu a jejich přenos po něm jsou šifrována, koncová zařízení pak opatřena mnohdy vícestupňovou ochranou. K přihlášení do účtů obsahujících citlivá data je mnohdy vyžadována dvoufázová autentizace, platby platební kartou po internetu si zase v současné době vyžadují opsání vygenerovaného SMS kódu z uživatelského telefonu.

Jako další stupeň ochrany se v posledním desetiletí začíná dostávat do spotřebitelského průmyslu i biometrické ověřování totožnosti uživatele. Biometrie je obor zabývající se měřením a vyhodnocováním biologických charakteristik lidí, nejčastěji dnes na základě otisku prstu, skenu oční duhovky, nebo rozpoznávání obličeje.

Právě výše zmíněné metody autentizace v posledních letech radikálně mění způsob, jakým používáme svá mobilní zařízení. Umožňují nám přihlášení do telefonu, následně pak i do aplikací v něm, provést pouhým přiložením prstu na senzor, nebo pohledem do kamery umístěné na přední straně přístroje. Tento způsob umožňuje radikálně snížit riziko přístupu cizí osoby k datům uživatele oproti použití běžného hesla či pinu.

Většina mobilních zařízení, zejména využívající operační systém Android, využívá pro autentizaci uživatele na základě obličeje pro sběr dat RGB kameru, k vyhodnocení pak umělou inteligenci. Oproti tomu, systém od společnosti Apple, nazvaný Face ID využívá pro sběr dat projektor, který oblast obličeje pokryje sítí bodů infračerveného světla, a kameru citlivou na infračervené světlo, která výsledný obraz vyhodnotí a porovná s lokálně uloženou předlohou, čímž se dosahuje znatelně vyšší bezpečnosti (díky snímání geometrických informací obličeje oproti pouhé barevné informaci) a nezávislosti na okolním osvětlení než v případě RGB kamery.

Technologie, kterou si společnost Apple vybrala, je používána ve velké škále 3D skenerů, sloužících ke skenování menších či větších objektů. Jedním z těchto skenerů je také běžně dostupný senzor Microsoft Kinect V2, původně vyvinut k ovládání herní konzole Xbox One, od společnosti Microsoft. Pro jeho velice nízkou cenu, uspokojivé přesnosti a otevřenému SDK (sada vývojových nástrojů umožňujících ovládat senzor pomocí programu) je použitelný pro autentizaci.

Zbytek práce je rozdělen následovně. V kapitole 2 je věnována pozornost metodám 3D rekonstrukce obličeje za pomoci hloubkových map. Kapitola 3 je zaměřena na průzkum existujících řešení a algoritmů pro co nejvýstižnější a zároveň úsporný popis obličeje. Možnostem a získávání základních obrazových dat ze senzoru Microsoft Kinect V2 je věnován prostor v kapitole 4. Kapitola 5 se zabývá realizací obslužného programu určenému pro sběr a 3D rekonstrukci dat.

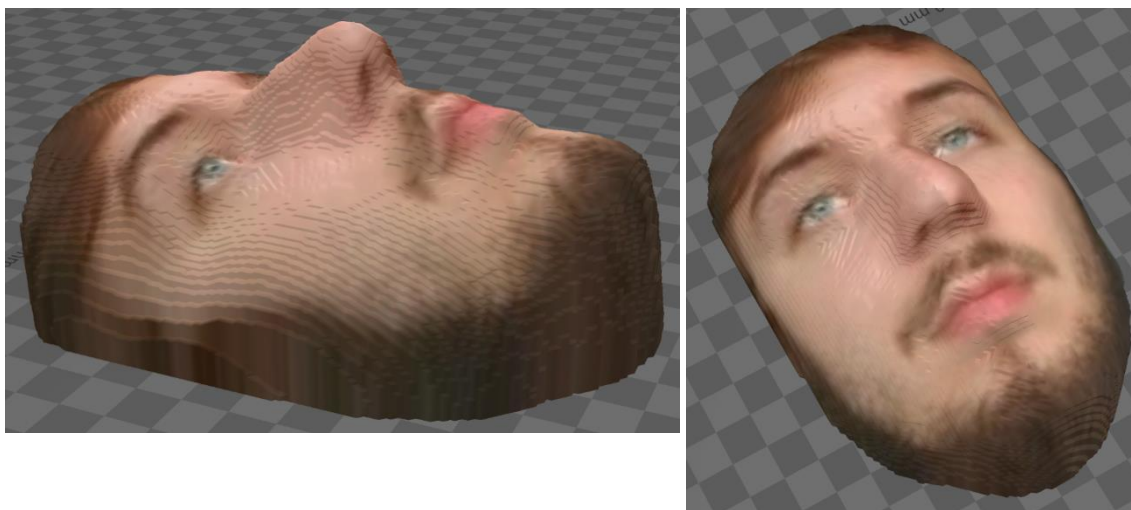
Kapitola 6 popisuje způsob, jakým jsou extrahovány popisné body skenovaného obličeje. O samotném porovnávání popisných bodů mezi jednotlivými uživateli pojednává kapitola 7. Kapitola 8 popisuje testovací sadu obličejů, která byla nasbírána za účelem ověření funkčnosti navržených algoritmů, součástí této kapitoly jsou také podmínky pro správnou funkčnost navrženého algoritmu. Kapitola 9 je věnována úvaze o použitelnosti navrženého systému v praxi.

2 METODY 3D REKONSTRUKCE OBLIČEJE

3D rekonstrukce je proces získávání prostorového tvaru obličeje z jednoho či více RGB snímků, hloubkové mapy, nebo RGB-D snímků za použití strukturovaného světla v blízkosti infračerveného spektra, což je kombinace předchozích dvou. 3D rekonstrukce je významným krokem pro zvýšení přesnosti a robustnosti autentifikace či identifikace osob. Hlavním důvodem k rekonstrukci je, že 3D model nese, oproti 2D snímku, geometrické údaje o obličeji, které jsou velice důležitou informací.

Rekonstrukce pouze pomocí RGB snímků je možná, avšak je nutné zachovat příznivé světelné podmínky, což v nějakých případech není možné. Zpracování takových snímků je také výpočetně náročné a o celý proces se starají neuronové sítě. Ačkoliv je tedy takový přístup zajiště možný a má své využití, pro účely identifikace osob je jeho použití sporné. Na obr. 1 je uveden výsledek 3D rekonstrukce obličeje na základě jednoho RGB snímku algoritmem navrženým v [1]. Jak lze vidět, rekonstrukce zachycuje pouze základní obličejové tvary, které byly navrženy neuronovou sítí. Z tohoto důvodu je tato metoda pro účely identifikace osob nevhodná.

Tyto nedostatky zvládne částečně řešit stereoskopické snímání, které dokáže lépe vypočítat geometrii obličeje díky vzájemnému posunu dvou kamer. Stále však přetrvává závislost na dobrých světelných podmínkách.



Obr. 1: 3D Rekonstrukce obličeje z fotografie algoritmem navrženou v [1]

Z výše zmíněných důvodů je velice oblíbené použití snímačů citlivých na světlo infračerveného spektra. V posledních letech se taková zařízení stávají relativně dostupnými a umožňují velice velkou nezávislost na světelných podmínkách (oproti RGB snímačům dosahují skvělých výsledků při nízké intenzitě osvětlení, některé snímače jsou ale ovlivňovány přímým slunečním světlem).

Pro snímání se využívá jeden z následujících principů. Scéna je buď aktivně osvětlována světlem o vlnové délce blízké infračervenému spektru a snímač zachycuje jeho odraz od objektů. Druhým principem je snímání infračerveného světla vyzařovaného objekty, což používají například termovize. Tyto snímače následně vyhodnocují přijaté infračervené světlo a pro každý pixel tak vypočítávají vzdálenost objektu.

Nevýhodou těchto přístupů je nižší rozlišení oproti běžným RGB kamerám a také vyšší šum. Jejich cena je v této době stále podstatně vyšší, proto také nejsou natolik rozšířené.

Pokud je tedy možné získat ze snímku hloubková data, je také možné z tohoto mračna bodů rekonstruovat samotný objekt. Vzhledem k šumu, který je u těchto snímačů přítomen, je většinou nutné samotný výpočet provést z průměru více měření, nebo na výsledná data použít mediánový či Gaussův filtr. Je třeba ale dbát na to, aby nebyla filtrem odstraněna i užitečná data.

Jak lze vidět na obr. 2, za použití velice dostupného senzoru Microsoft Kinect V2 využívajícího aktivně osvětlenou scénu infračerveným světlem, je možno dosáhnout znatelně vyšších detailů. V kombinaci s nasnímanými RGB daty je vytvořen zajímavý podklad pro další zpracování obličejových dat.



Obr. 2: 3D rekonstrukce obličeje pomocí algoritmu Kinect Fusion [12]

3 POPIS OBLIČEJOVÝCH DAT

Aby nebylo nutné při každé autentifikaci procházet celou databází uživatelů a porovnávat mezi sebou celé obličejové rekonstrukce, je potřeba uložit do databáze pouze ta data, která jsou pro identifikaci nezbytně nutná (snížit jejich dimenzi) a zároveň jich uložit dostatečné množství na to, aby byl celý systém stále bezpečný. Zároveň je takto, společně se šifrováním samotné databáze, předcházeno nebezpečí zneužití dat k neoprávněné autentifikaci při jejich odcizení.

Data je možné z rekonstrukce získat obecně třemi způsoby [2].

- Holistický přístup (Holistic matching)
- Lokální přístup (Feature-based matching)
- Hybridní přístup (Hybrid matching)

V následujících podkapitolách jsou blíže popsány nejpoužívanější metody výše zmíněných přístupů. Je běžnou praxí, že se pro jeden systém využívá kombinace více metod za účelem větší robustnosti a bezpečnosti.

3.1 Holistický přístup

Prvním způsobem je holistický přístup, při kterém je obličej popisován jako celek, proto jsou i výstupní parametry vypočteny na základě všech dostupných obličejových dat. To v praxi nejjednodušeji znamená, že do databáze je uložena celá rekonstrukce, kterou je pak v případě identifikace snaha zarovnat s novou rekonstrukcí. Míru shody pak určuje nejmenší nalezená akumulovaná odchylka. Tento přístup je značně náchylný na částečné zakrytí obličeje (brýle, zakrytí rukou) což dle způsobu využití může či nemusí vadit [3].

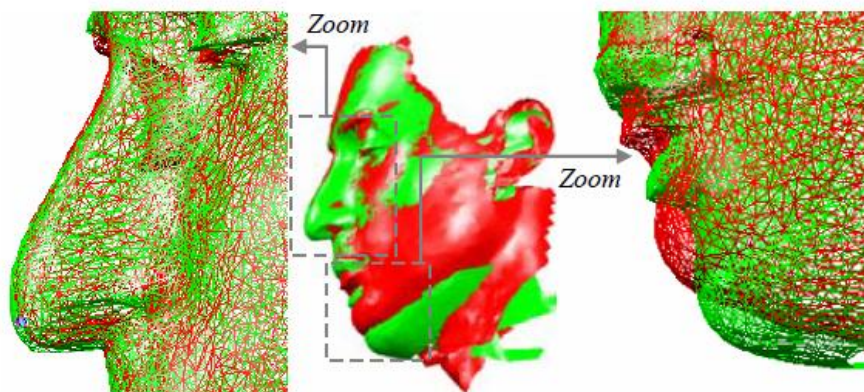
3.1.1 ICP (Iterative closest point)

Nejlogičtější metodou, která se nabízí při porovnávání dvou 3D modelů, je jejich precizní zarovnání a porovnání jejich odlišností. Metoda navržená v [3] rozděluje tento proces na dva významné kroky.

Prvním z nich je hrubé zarovnání (coarse alignment), které má za úkol tyto dva modely posunout do přibližně stejné pozice na základě několika významných bodů na obličeji. Tento krok ve výše zmíněné publikaci vyžaduje zásah operátora, který manuálně vybere dané body na obou porovnávaných modelech. Pro účely rozpoznávání obličejů toto ale žádoucí není, proto je třeba využít jednu z metod lokálního přístupu pro nalezení malého počtu bodů, které pak budou sloužit jako záchytné body pro hrubé zarovnání (například nos a oční koutky). Přesnost tohoto kroku dále určuje kvalitu celého systému identifikace.

Druhým krokem je iterativní algoritmus pro jemné zarovnání (fine alignment). Tento krok postupně snižuje odchylku sousedících bodů obou modelů metodou nejmenších čtverců až do bodu konvergence. Jak rychle a jestli bude tento krok konvergovat pouze k lokálnímu, nebo ke globálnímu minimu, závisí na přesnosti hrubého zarovnání. Na obr. 3 je uveden příklad vizualizace zarovnání dvou 3D modelů metodou ICP. Výsledek klasifikace je udán předem danou hranicí kumulované odchylky, pod kterou budou obličeje považovány za shodné.

Slabostí této metody je výpočetní náročnost při velké databázi uživatelů. Další nevýhodou je pak citlivost na změny výrazů, která lze částečně potlačit rozdělením obličeje do sekcí, které podléhají či nepodléhají změnám při výrazech v obličeji. Odchyly mezi modely v těchto sekcích se pak do výsledné kumulované odchylky přičítají s různými vahami.



Obr. 3: Vizualizace zarovnání 3D modelů pomocí ICP metody [3]

3.1.2 PCA (Principal component analysis)

Analýza hlavních komponent je jednou z nejrozšířenějších metod snižování rozměru dat hlavně díky své relativní jednoduchosti. Často je využívána pro rozpoznávání objektů, většinou kombinována se strojovým učením. Díky své schopnosti drasticky snížit rozměr dat bez potenciálně výrazné ztráty informace je tato metoda vhodná i pro rozpoznávání obličejů.

Metoda navržená v [4] využívá 3D model referenčního obličeje, který je aproximací zarovnan s testovaným obličejem a z jejich odlišností je pak vypočtena kovarianční matice, ze které je vyjmut vektor popisující testovaný obličej. Díky tomu, že jsou všechny testované obličeje porovnávány s identickým referenčním modelem (existuje tedy pro všechny stejný referenční bod), je možné porovnávat výsledné vektory i mezi sebou. Nebýt tohoto, bylo by nutné zarovnávat a porovnávat všechny modely mezi sebou, což by bylo výpočetně velice náročné. Na obr. 4 je uveden příklad zarovnání testovaného obličeje s referenčním.

Nevýhodou této metody (stejně jako ostatních metod holistického přístupu) je ne příliš velká robustnost oproti zakrytí obličeje (například brýle, nebo takové natočení hlavy, které skryje výhled na zbytek obličeje). Robustnost proti různým výrazům v tváři také není příliš velká, proto má tato metoda daleko větší úspěch v odvětvích, kde uživatel maximálně kooperuje s programem. Metoda, která se s těmito vlivy snaží bojovat byla navržena v [5]. Nespornou výhodou této metody je její rychlost.



Obr. 4: Příklad zarovnání obličeje s referenčním obličejem pro klasifikaci metodou PCA [4]

3.2 Lokální přístup

Metody založené na lokálním přístupu popisují, oproti metodám holistickým, pouze ty vlastnosti obličejových dat, které jsou nějakým způsobem charakteristické. Díky této vlastnosti jsou daleko robustnější oproti změnám výrazu či zakrytí části obličeje. Z důvodu vyšší výpočetní náročnosti při spolehlivém určování lokálních charakteristik, je nutné volit pouze ty, které mají maximální vypovídající hodnotu a zároveň nejsou náchylné na změnu polohy při různých výrazech v obličeji, nebo například stárnutí.

Vlastnostmi, které se v praxi často pro tyto účely používají, mohou být například skupiny specifických bodů a jejich geometrické uspořádání, nebo také charakteristické obličejové křivky či plochy.

Metody lokálního přístupu fungují ve dvou velice obecných krocích. Prvním z nich je samotná detekce význačných bodů (keypoint detection), druhým pak jejich interpretace (feature description) [2]. Teprve pak mohou být porovnávány mezi obličejí. V následujících podkapitolách jsou blíže popsány nejpoužívanější metody lokálního přístupu.

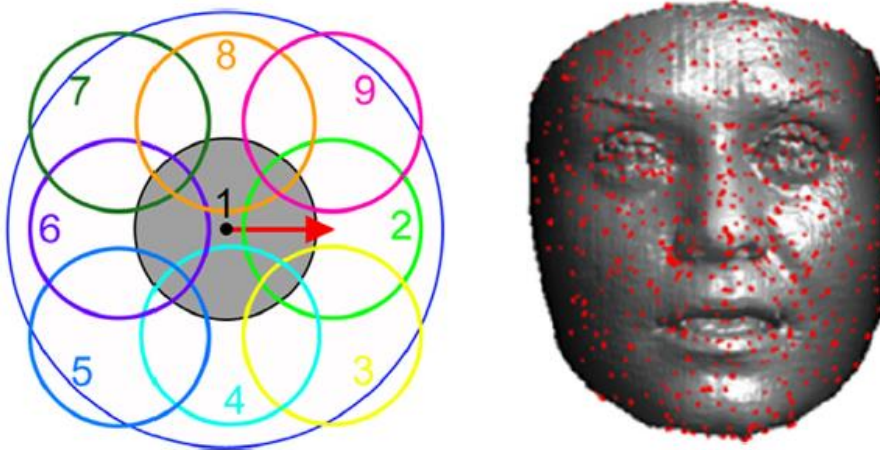
3.2.1 Metody klíčových bodů (Keypoint-based methods)

Tyto metody hledají definované význačné body v obličejových datech. Předem definované body jsou detekovány na základě geometrických vlastností povrchu. Přestože tyto metody mohou odstranit vliv zakrytí části obličeje nebo chybějícími částmi (například z důvodu, že uživatel není natočen přímo proti senzoru), je jejich výpočetní náročnost velice vysoká. Z tohoto důvodu je velice důležité pečlivě volit body, které jsou pro daný účel efektivní.

3.2.1.1 Síťové metody (Mesh-based methods)

Metoda navržená v [6] je založena na detekování význačných bodů na základě lokálních maxim či minim v zakřivení obličeje. Často se pro lokalizaci těchto bodů využívá modifikovaná metoda SIFT (Scale Invariant Feature Transform), která se hojně využívá i při zpracovávání 2D obrazových dat. Pomocí rozdílu obličejových dat, na které byl použit Gaussův filtr o různých silách jsou vypočteny význačné body. K těmto bodům je následně přiřazena jejich orientace a jejich okolí je rozděleno a popsáno histogramy zachycujícími místní zakřivení a úhly.

Při samotné identifikaci se pak porovnávají dvě sady bodů s jejich přidruženými informacemi. Počet identických bodů pak určuje, zda se jedná o obličej stejné osoby či nikoliv.



Obr. 5: Detekované body síťovou metodou (vpravo) a rozložení jejich okolí pro účely popisu (vlevo) [13]

3.2.1.2 Metody význačných bodů (Landmark-based methods)

Tyto metody pracují s body vycházejícími z anatomických studií obličeje. Těmito body může být například nos, koutky očí, obočí či ústa. Na základě jejich geometrického uspořádání je posléze provedena identifikace.

Očividnou nevýhodou těchto metod je skutečnost, že existuje poměrně málo anatomicky význačných bodů, z toho důvodu je nutné provést jejich extrakci s vysokou přesností, aby bylo dosaženo vysoké spolehlivosti. Existuje mnoho metod pro extrakci, v [7] je vypočteno zakřivení obličeje pomocí druhých derivací hloubkových map. Jednoduchým prahováním se pak izolují místa s největším zakřivením, která odpovídají špičce nosu a vnitřním koutkům očí.

V poslední fázi je sestaven trojúhelník s vrcholy v nalezených bodech, který následně slouží pro porovnávání obličejových dat ve fázi rozpoznávání. Na obr. 6 a 7 je znázorněn proces segmentace obličeje za účelem nalezení zmíněných bodů a vynesení trojúhelníku, který je spojuje.

Výhodou této metody je robustnost oproti zakrytí úst nebo čela. Naopak brýle by mohly představovat problém. Cena své relativní jednoduchosti je pak ale zaplácena nutností precizní extrakce bodů, jinak by nemohlo být dosaženo dostatečné přesnosti. Jak je také uvedeno v [7], tato metoda lze jednoduše využít pro samotnou detekci obličejů v hloubkových mapách, kdy dokáže detekovat i více osob najednou.



Obr. 6: Ukázka extrakce bodů s největší, zakřivením pomocí prahování [7]



Obr. 7: Vizualizace unikátního trojúhelníku pro popis obličeje [7]

3.2.2 Křivkové metody (Curve-based methods)

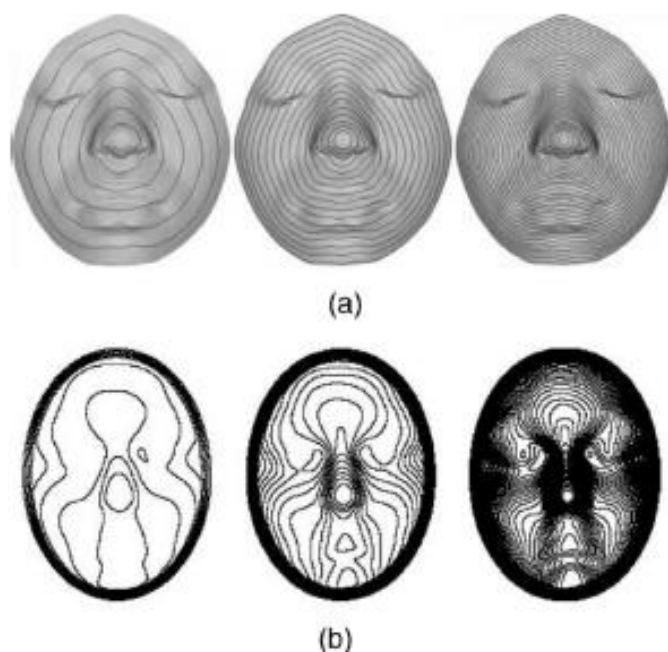
Tyto metody využívají pro identifikaci obličeje soubor charakteristických křivek, které nesou důležité informace o tvaru obličeje. V porovnání s metodami klíčových bodů jsou méně náchylné na změny výrazu. Pro docílení lepších výsledků lze křivkám v různých polohách přiřazovat vyšší váhy (často např. špička nosu). Křivkové metody mohou být rozděleny do dvou hlavních kategorií [8]

3.2.2.1 Konturové metody (Contour-based methods)

Metody jsou založeny na „konturách“, neboli uzavřených křivkách o různých délkách, které se mezi sebou neprotínají. Kontury spojují místa se stejnou hloubkou (stejně jako vrstevnice), čehož je docíleno proložením plochy ve vodorovné ose skrz obličej. Tento „hloubkový způsob“ je využit v [9].

Druhou možností jsou křivky, které zachovávají určitou vzdálenost od kotevního bodu („způsob konstantní vzdálenosti“). Metoda použita v [10] detekuje špičku nosu a následně počítá všech bodů vzdálenost po povrchu obličeje od tohoto místa. Spojením bodů se stejnou vzdáleností vznikají křivky, které jsou, dle uvedeného zdroje, velice robustní proti změnám výrazu obličeje.

Kontury zachycují obecně více dat, než například metody kapitoly 3.2.1.23.2.1.2, na druhou stranu však nezachycují kompletní okolí nosu, které má největší vypovídající hodnotu. Na obr. 8 jsou vyobrazeny vizualizace jednotlivých metod.



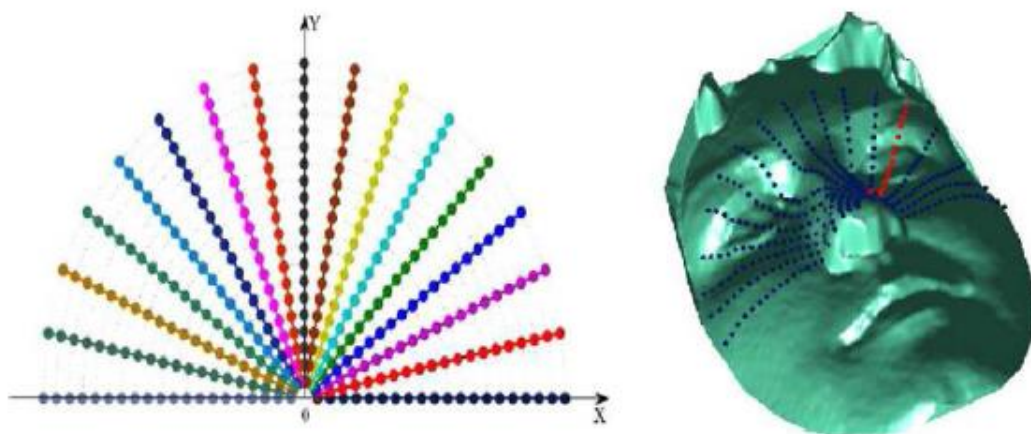
Obr. 8: Vizualizace konturových metod hloubkovým způsobem (a), způsobem konstantní vzdálenosti (b) [2]

3.2.2.2 Profilové metody (Profile-based methods)

Profilové křivky, na rozdíl od konturových zmiňovaných v kapitole 3.2.2.1, nejsou uzavřené. Mají tedy počáteční a koncový bod. Počáteční bod bývá volen na špičce nosu, koncové pak na okrajích obličeje.

Po detekci referenčního bodu jsou tedy vypočteny křivky do všech směrů ke okraji obličeje. Z důvodu snižování výpočetní náročnosti je dobré rozumně volit rozestup mezi křivkami, nebo také brát v úvahu pouze ty směry, kde se očekává menší náchyllost na změny výrazu. Díky směrovosti se jedná o zajímavou metodu oproti konturám, protože lze potlačit vliv zakrytí části obličeje, kde by si konturové křivky nevedly tak dobře.

Metoda navržená v [11] představuje binární masku upřednostňující pouze křivky vrchní poloviny obličeje, jak je znázorněno na obr. 9.



Obr. 9: vizualizace profilové metody (vpravo) s binární maskou (vlevo) [11]

3.3 Hybridní přístup

Hybridní přístup je pouhou kombinací lokálních a holistických metod. Můžeme tak například pomocí lokálních metod popisovat 3D model obličeje a na jeho texturu využít jednu z globálních metod. Teoreticky se tím stává systém spolehlivější a robustnější oproti různým vlivům.

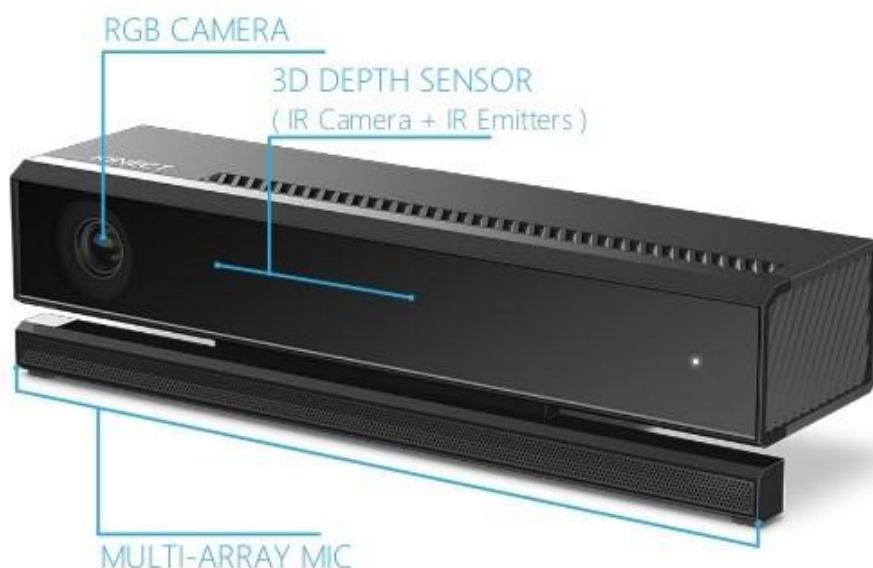
Kombinace těchto metod silně závisí na zamýšleném použití celého systému, jak již bylo zmíněno, každá z metod má své silné stránky, má však i své slabiny. Obecné výhody holistických metod jsou v rychlosti a v hledání obecných podobností v obličejových datech. Naopak lokální metody jsou obecně robustnější proti překrytí obličeje, natočení či výrazům. To však ale za cenu vyšší výpočetní náročnosti a důrazu na velice přesnou extrakci bodů či křivek.

4 MICROSOFT KINECT V2

Tato kapitola je věnována představení senzoru Microsoft Kinect V2, zejména po jeho hardwarové stránce, jsou představeny jeho senzory a jejich přesnosti. Zřetel je brán zejména na popis formátu výstupních dat z jednotlivých senzorů, což dále určuje způsob, jakým s nimi bude nakládáno.

Zařízení Microsoft Kinect V2 bylo vyvinuto jako ovladač k herní konzoli Xbox One. Díky své kombinaci senzorů umožňuje uživatelům ovládat herní konzoli za pomoci přirozených gest, nebo mluvených příkazů. Uživatelům je zpracováním dat z jednotlivých snímačů určeno až 25 bodů na těle. Jedná se především o všechny velké klouby, pro účely rozeznávání gest je snímán také palec ruky oproti zbývajícím prstům. Zařízení je takto schopno snímat až 6 uživatelů v prostoru najednou. Microsoft Kinect V2 funguje za pomoci spojení dat z několika snímačů. Obsahuje RGB kameru, IR kameru, IR projektor a pole mikrofonů. Použití pro herní konzoli, jak se později ukázalo, se bohužel nesešlo s velkým ohlasem u svých uživatelů, proto bylo zařízení později staženo z trhu.

Výhody tohoto zařízení byly později objeveny nezávislými vývojáři a výzkumníky, kteří si jej oblíbili pro jeho vlastnosti a nízkou cenu. S pomocí sady vývojových nástrojů (SDK), které společnost Microsoft uvolnila, je možno senzor ovládat a získávat data z jednotlivých snímačů v nezpracované podobě (RGB obraz, obraz infračerveného spektra), nebo jiná zajímavě zpracovaná data (hloubková mapa). Fúzí těchto datových toků jsou pak vývojové nástroje schopny vypočítat polohu uživatele, rozpoznat obličej a výraz v něm.



Obr. 10: Rozložení snímačů v senzoru Microsoft Kinect V2 [14]

4.1 RGB Kamera

V senzoru je použit snímač s rozlišením 1920x1080 pixelů (FullHD) schopný získávat snímky 30x za sekundu. Zorné pole snímače je horizontálních 84° a vertikálních 54°. S tímto zorným polem je senzor schopen shlédnout prostor 90x50 cm na vzdálenosti 50 cm a až 9x5m na vzdálenosti 5 m.

Data ze snímače je možno získat asynchronně z barevného datového proudu v reálném čase.

4.2 Infračervená kamera

V senzoru je použit snímač s rozlišením 512x424 pixelů schopný získávat snímky 30x za sekundu. Zorné pole snímače je horizontálních 70° a vertikálních 60°. Tím je snímač schopen obzoru 70x57 cm na vzdálenosti 50 cm a až 7x5.7 m na vzdálenosti 5 m.

Scéna je aktivně osvětlována infračerveným projektorem modulovaným na frekvenci okolo 100MHz.

Data ze snímače můžeme získat asynchronně v reálném čase v čisté podobě z infračerveného datového proudu, nebo zpracované v podobě hloubkové mapy z hloubkového datového proudu.

4.3 Datové proudy

Jak již bylo avizováno v předchozím textu, sada vývojových nástrojů pro Kinect V2 nám umožňuje asynchronně získávat data ze svých snímačů. Kromě čistých nezpracovaných dat (RGB, IR, audio), je možné získat zpracovaná data (Hloubková mapa, pozice uživatele). Pro přístup k jednotlivým datovým proudům v reálném čase je nejprve nutné samotný senzor inicializovat, definovat o jaké datové proudy máme zájem a vytvořit asynchronní metodu, která bude zavolána, když budou data k dispozici. Na obr. 11 je uveden příklad kódu využívajícího sadu vývojových nástrojů k získání dat z jednotlivých snímačů.

```
_sensor = KinectSensor.Default();

if (_sensor != null)
{
    _sensor.Open();

    _reader = _sensor.OpenMultiSourceFrameReader(FrameSourceTypes.Color | FrameSourceTypes.Depth | FrameSourceTypes.Infrared);

    _reader.MultiSourceFrameArrived += Reader_MultiSourceFrameArrived;
}
```

Obr. 11: Příklad inicializace senzoru Microsoft Kinect V2

1 reference | 0 changes | 0 authors, 0 changes

```
void Reader_MultiSourceFrameArrived(object sender, MultiSourceFrameArrivedEventArgs e)
{
    var reference = e.FrameReference.AcquireFrame();

    using(var frame = reference.ColorFrameReference.AcquireFrame())
    {
        if(frame != null)
        {
            colorImage.Source = ToBitmap(frame);
        }
    }

    using (var frame = reference.InfraredFrameReference.AcquireFrame())
    {
        if (frame != null)
        {
            infraredImage.Source = ToBitmap(frame);
        }
    }

    using (var frame = reference.DepthFrameReference.AcquireFrame())
    {
        if (frame != null)
        {
            depthImage.Source = ToBitmap(frame);
        }
    }
}
```

Obr. 12: Metody přejímající data z jednotlivých datových toků asynchronně

Snímky z datových proudů je pak nutné dále zpracovat účely aplikace, nebo pouhého zobrazení na obrazovku uživatele. V následující podkapitole je představen hloubkový datový tok, který pochází z infračervené kamery a je dále matematicky zpracováván.

4.3.1 Hloubkový datový proud

Výpočet hloubkové mapy je v případě senzoru Microsoft Kinect V2 prováděn na základě nepřímé Time-of-flight (indirect ToF) metody. Tato metoda využívá známé rychlosti světla pro výpočet vzdálenosti mezi snímačem a objektem. Jak bylo zmíněno na začátku kapitoly 2., scéna před senzorem je aktivně osvětlována infračerveným projektorem, jehož výstup je modulován frekvencí okolo 100MHz. Toto vyslané světlo se po odražení od objektu dostává zpět k senzoru na infračervenou kameru, která jej zachytí a jsou porovnány fáze světla vyslaného a přijmutého. Výsledná vzdálenost je pak vypočtena pomocí vzorce:

$$d = \frac{\Delta\varphi}{4\pi f} \cdot c$$

C ... rychlost světla

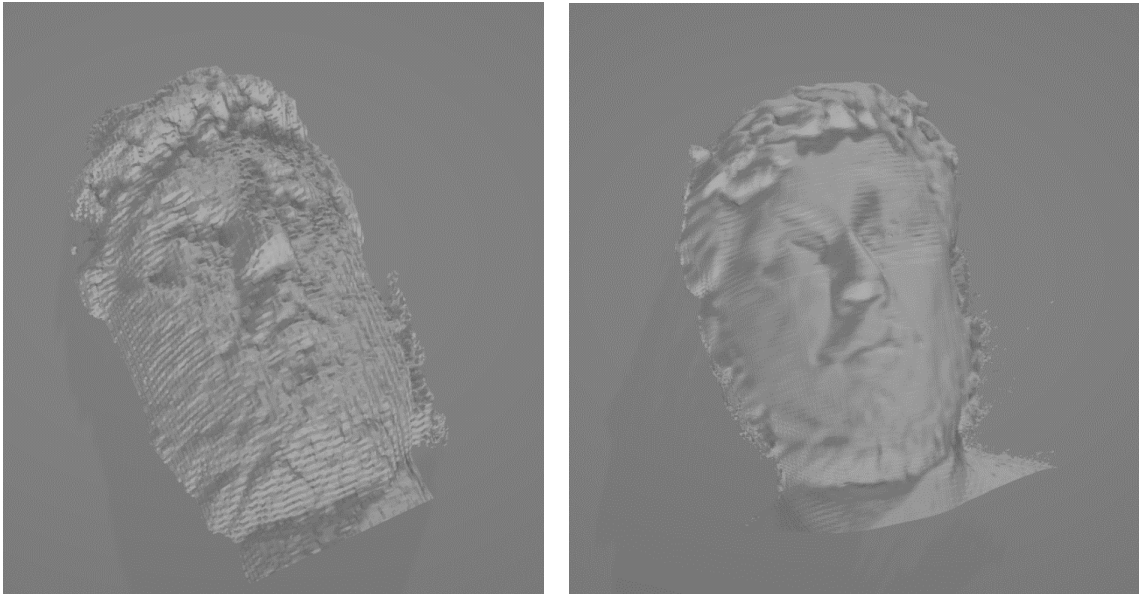
$\Delta\varphi$... fázový rozdíl

f ... modulační frekvence

Vypočtená data jsou pak odeslána do programu jako matice vzdáleností v rozmezí 50 cm až 450 cm, což jsou hardwarové limity tohoto senzoru.

4.4 Kinect Fusion

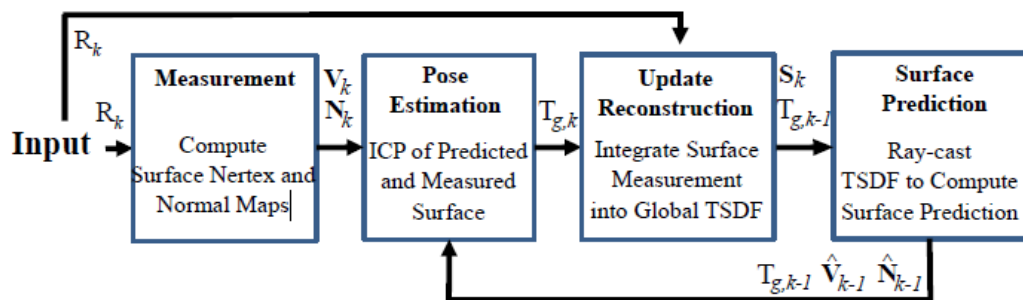
Pro potřeby rozpoznávání obličejových dat nestačí zpracovat pouze snímek z hloubkového datového proudu. Takový snímek je značně nedokonalý, obsahuje mnoho šumu a místy mohou i naměřená data zcela chybět. Všechny tyto nedostatky by velice ovlivňovaly přesnost takového systému autentifikace. Na obr. 13 příklad obličejových dat z jednoho snímku. Jak je vidno, sken zachycuje pouze základní obličejové rysy. Je nutno podotknout, že v okolí obličeje se nacházelo mnoho šumu, který musel být manuálně odstraněn pro účely tohoto příkladu.



Obr. 13: Příklad 3D rekonstrukce obličeje z jednoho hloubkového snímku (vlevo), z více snímků (vpravo)

Z tohoto důvodu je nutné použít více snímků v řadě a výsledná data spojit do jednoho výsledného modelu. Microsoft ve svém balíčku vývojových nástrojů pro senzor Kinect V2 přikládá algoritmy, které umožňují 3D rekonstrukci objektů v reálném čase. Tento nástroj umožňuje přidat každou získanou hloubkovou mapu do jedné společné 3D rekonstrukce objektu i v případě, že je senzor v pohybu. Je tak možno nasnímat objekt z více stran a zvyšovat jeho výstupní rozlišení. Na obrázku výše je uveden příklad modelu vytvořeného z přibližně 60 snímků.

Takový model již může být použit pro účely identifikace. Další možností algoritmu Kinect Fusion je přiřadit výslednému modelu data z RGB kamery, která by teoreticky zvýšila přesnost celého systému autentifikace. V následujících podkapitolách je vysvětleno, jak algoritmus Kinect Fusion funguje [12].



Obr. 14: Blokové schéma procesu rekonstrukce scény pomocí algoritmu Kinect Fusion [12]

4.4.1 Snímání prostředí

Počáteční fáze, kde je z dat pocházejících z hloubkového datového proudu R_k vygenerováno mračno bodů V_k a z nich vypočteny normály jednotlivých ploch N_k . Na vstupní data R_k je použit bilaterální filtr pro snížení šumu při zachování detailů. Tato výstupní data jsou dále zpracována v následujícím kroku

4.4.2 Vyhodnocení polohy senzoru

Síla tohoto nástroje spočívá mimo jiné i ve schopnosti mapovat prostor i za pohybu samotného senzoru. Díky obsaženému akcelerometru a předchozím hloubkovým mapám je možné vypočíst novou polohu senzoru $T_{g,k}$ pro každou příchozí hloubkovou mapu. Díky rychlosti, s jakou jsou hloubkové mapy generovány (30 Hz) je také možno předpokládat, že relativní změna polohy mezi po sobě jdoucími snímky bude minimální.

4.4.3 Aktualizace mračna bodů

Na základě polohy senzoru $T_{g,k}$ vypočtené v předchozím kroku a za použití nejaktuálnější hloubkové mapy R_k jsou do výsledného mračna bodů přidána nová data. S každým snímkem se tak zvyšuje rozlišení výsledného modelu, snižuje se výsledný šum, nebo je objekt nasnímán z jiného úhlu, pokud se senzor pohybuje.

4.4.4 Výpočet výsledného modelu

S nejaktuálnějším mračnem bodů je již možno vytvořit plochy výsledného modelu a například ho vykreslit na obrazovku. Tento model je pak použit pro výpočet polohy senzoru v dalším snímku a při další iteraci je opět doplněn o další naměřená data. Ve chvíli, kdy je model dostatečně přesný, je možné rekonstrukci zastavit a uvolnit tak systémové prostředky pro další operace. V této fázi lze k výslednému modelu přiřadit také obrazový výstup z RGB kamery, který může v následné autentifikaci výrazně pomoci.

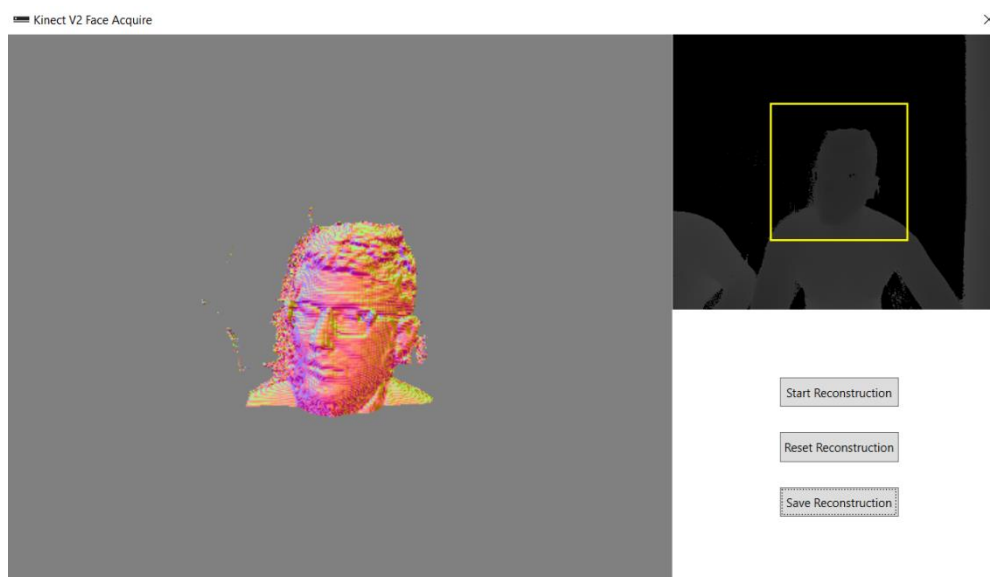
5 OBSLUŽNÝ PROGRAM

Obslužný program senzoru Microsoft Kinect V2 pro sběr dat a jejich exportování je realizován v programovacím jazyce C#. Pro tvorbu grafického rozhraní je využita knihovna WPF.

Program ihned po spuštění spustí infračervený projektor a kameru pro hloubkové snímání prostoru před sebou. Na obrazovce je zobrazena hloubková mapa s vizualizací snímaného prostoru a je po uživateli vyžadována kontrola správné pozice před senzorem. Ve chvíli, kdy je uživatel před senzorem ve správné pozici, je třeba stisknout tlačítko pro start 3D rekonstrukce obličeje.

Samotná rekonstrukce je složením předem definovaného počtu hloubkových snímků, které lze taktéž nastavit v grafickém rozhraní obslužného programu. Celý proces rekonstrukce je zobrazen v prostřední části okna programu a slouží k manuální kontrole správnosti před uložením obličejových dat do databáze. Pokud je uživatel spokojen s kvalitou rekonstrukce, je třeba stisknout tlačítko „Uložit rekonstrukci“.

V následujících podkapitolách jsou v bodech popsány nejdůležitější body procesu sběru dat v obslužném programu.



Obr. 15: Ukázka obslužného programu

5.1 Počáteční inicializace senzoru

Při startu obslužného programu je inicializován datový proud RGB kamery a hloubkový datový proud infračervené kamery. Přesnost senzoru je dále konfigurována. Maximální hloubka, se kterou bude program pracovat je 1 metr. Hodnota rozlišení senzoru na 768 voxelů (bodů v 3D prostoru) na metr, což znamená, že jednomu metru bude v hloubkovém snímku odpovídat 768 bodů. Následně je omezen maximální počet voxelů v každé ose tak, aby byl dostatečný prostor na obličej, ale zároveň nebylo snímáno případné okolí.

Tato nastavení mají několik důvodů. Omezením maximální hloubky značně snižujeme náročnost samotné 3D rekonstrukce, nastavením rozlišení senzoru zaručujeme maximální dostupnou rozlišovací schopnost, která je v tomto případě okolo 1,3mm.

5.2 Přejímání dat z datových proudů

Data z datových proudů jsou přejímána asynchronní funkcí, která vždy přijme dvojici barevného a hloubkového snímku najednou. Příchozí data jsou následně validována a zkopírována do pole pro následné zpracování. Pokud jsou data v pořádku, je nastaven příznak, který signalizuje, že jsou připravena pro rekonstrukci. Senzor Microsoft Kinect V2 je schopný zasílat data až 30x za sekundu.

5.3 Zpracování přijatých dat

V pozadí programu běží vlákno, které čeká na nastavení příznaků signalizujících připravená data. Jakmile je příznak nastaven, spustí se rutina, která spolupracuje s knihovnou Kinect Fusion pro úvodní vytvoření či přidání dat do výsledné rekonstrukce.

Nezbytnou součástí zpracování příchozích dat je přepočítání dat z RGB kamery do souřadnicového systému infračervené kamery, tento přepočítání je nutný z důvodu fyzického posunutí snímačů na senzoru.

Posledním krokem je předání dat zpět programu pro zobrazení průběhu rekonstrukce na obrazovku uživatele.

5.4 Uložení 3D rekonstrukce

Po úspěšné rekonstrukci a stisknutí tlačítka „Uložit rekonstrukci“ je vyvolána obsluha uložení výsledných dat ve formátu PLY (polygon file format) nebo CSV (comma-separated values). Datový formát PLY byl navržen na ukládání dat získaných ze 3D skenerů a podporuje uložení mnoha podpůrných informací, které dále popisují 3D model (barva, normálové vektory, atd.). Je také podporován většinou programů pro vizualizaci 3D dat. CSV datový formát v našem případě slouží jako nejjednodušší varianta pro ukládání a následné zpracovávání dat v rámci skriptu pro identifikaci obličejů.

5.5 Automatizace sběru dat

Pro zvýšení komfortnosti celého procesu by bylo dobré automaticky detekovat obličej před senzorem. K tomuto se nabízí použití metody představené v kapitole 3.2.1.2. Detekce by probíhala bez omezení maximální zpracovávané hloubky z infračervené kamery.

Ve chvíli, kdy by algoritmus detekoval v obraze lidský obličej, byla by maximální rekonstruovaná hloubka a ostatní parametry automaticky nastaveny tak, aby zahrnuly pouze detekovaný obličej. Tento algoritmus není tak užitečný při úvodním sběru dat pro uložení uživatele do databáze, je však velice příhodný pro použití ve stádiu sběru dat pro identifikaci uživatele.

Dalším zlepšením procesu je vytvoření databáze, kam se budou data automaticky ukládat ihned po nasnímání bez nutnosti manuálního uložení tlačítkem „Uložit rekonstrukci“.

6 EXTRAKCE POPISNÝCH BODŮ

Pro účely této práce byl algoritmus pro zpracování obličejových dat vytvořen v programovacím prostředí MATLAB. Výhodou tohoto přístupu je velmi jednoduchá práce s velkými objemy dat, které představují skenované obličeje. Dále je možné velice rychle krokovat programem a zobrazit si v každé chvíli veškeré proměnné v přehledných tabulkách, či je vykreslit do grafů. MATLAB je navíc skriptovacím jazykem, není tedy nutný mnohdy zdoluhavý překlad programu po každé drobné změně. Všechny tyto faktory ukazují na tuto variantu jako na ideální pro rychlé prototypování celého řetězce od filtrace vstupních dat až po porovnávání výstupních dat mezi jednotlivými obličeji.

Naopak značnou nevýhodou tohoto přístupu je nutnost následně algoritmus přepsat do jiného jazyka a sjednotit ho tak s algoritmem pro akvizici dat, pokud má být jakýmkoliv způsobem komerčně využíván.

Vstupními daty pro tento algoritmus je soubor ve formátu CSV vygenerovaný v předchozím kroku. Máme tedy pro každý bod v prostoru dostupnou jeho pozici, barvu i normálový vektor.

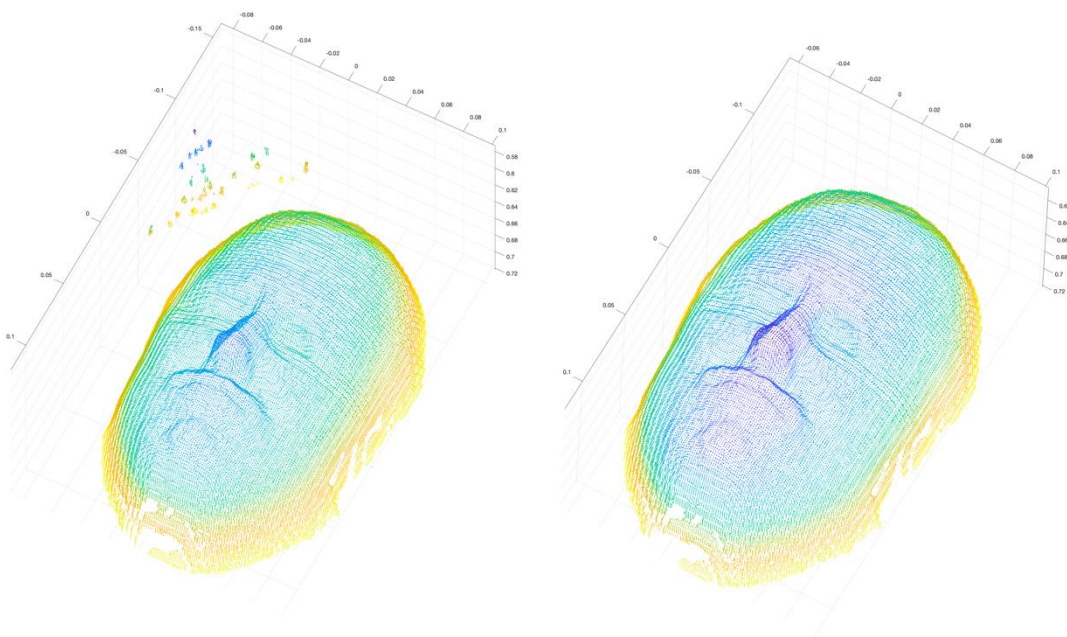
6.1 Předzpracování vstupních dat

Výstupní data z akvizičního algoritmu jsou mnohdy zatížena chybami ve formě šumu mimo oblast samotného obličeje. Před extrakcí obličejových dat je nezbytně nutné tento šum odstranit, jinak hrozí, že algoritmus nebude správně fungovat. Po odstranění šumu je následně vhodné segmentovat pouze samotný obličej a odstranit tak například vlasy, které nemají žádnou popisnou hodnotu.

6.1.1 Počáteční filtrace

Pro filtraci mračen bodů nabízí MATLAB velice příhodnou funkci *pcdenoise* která je součástí balíčku *Computer Vision Toolbox*. Vstupní proměnnou této funkce je kromě samotného mračna bodů také počet nejbližších bodů, se kterými bude každý bod porovnáván. Při filtraci budou vyřazeny ty body, které jsou od svých nejbližších bodů vzdáleny více, než je průměrná vzdálenost pro celé mračno.

Toto v praxi znamená, že malé shluky bodů, které jsou vzdáleny od obličeje několik centimetrů a jejichž počet je menší než námi definovaná hranice, budou odstraněny. Na obr. 17 jsou vidět vstupní data před a po filtraci.



Obr 16: Vizualizace vstupních dat před filtrací (vlevo) a po filtraci (vpravo)

6.1.2 Segmentace obličeje

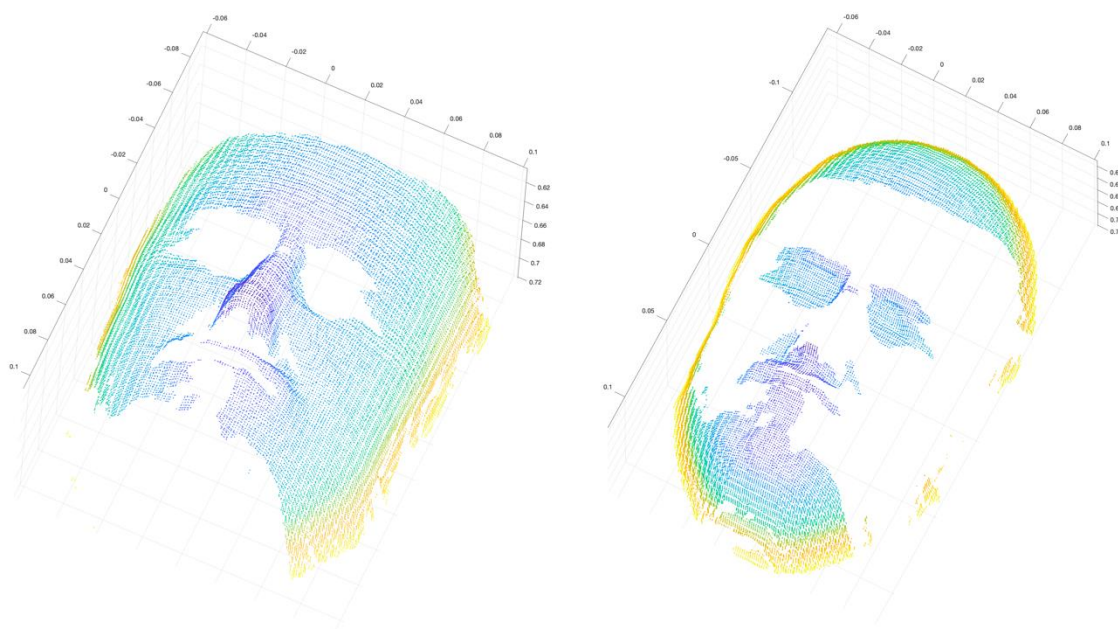
Vyfiltrovaná data od šumu je již možno segmentovat. Segmentace by v tomto případě měla docílit oddělení dat, která jsou pro účely rozpoznávání obličeje zajímavá od dat, jejichž využití nemá smysl. To, co algoritmus v tuto chvíli zachová, bude pouze pokožka, která bude v následující části rozdělena do dvou zón. Vlasy, fousy či pokrývky hlavy budou odstraněny. V průběhu vývoje byly pro tento účel použity celkem 3 postupy.

Nejprve bylo vyzkoušeno prahování pouze na základě barvy. To se ze zjevných důvodů neosvědčilo, pleť má totiž různou barvu jak v závislosti na ročním období, tak při různých podmínkách osvětlení. I při dodržení striktních podmínek osvětlení může nastat problém například se světlejšími vlasy.

Dalším způsobem bylo prahování na základě světlosti povrchu, což se neosvědčilo z velice podobných důvodů, jako předchozí varianta.

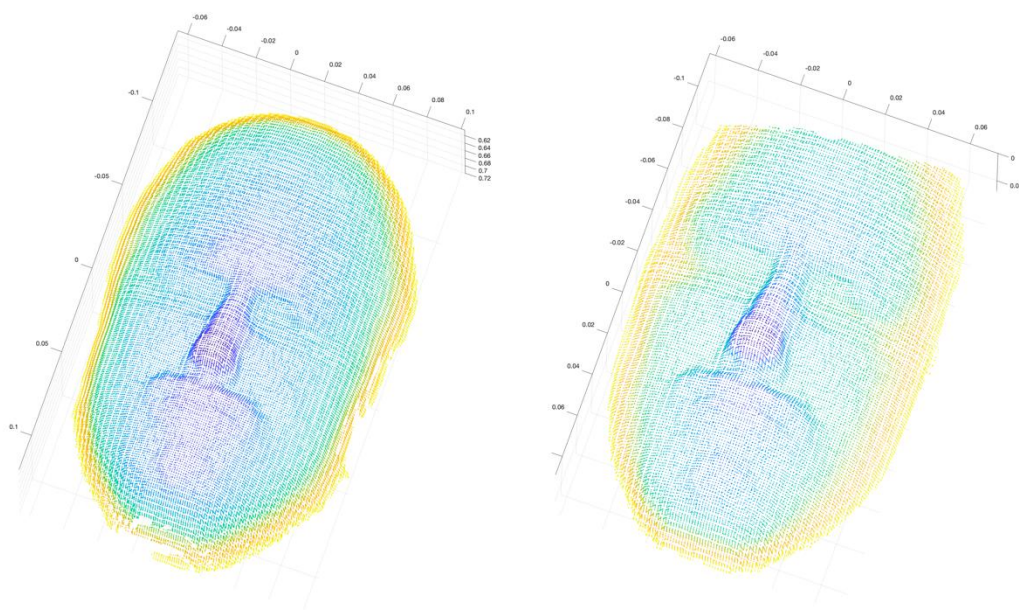
Způsob, který se osvědčil a vyniká svou robustností a relativní jednoduchostí na implementaci je využití funkce *kmeans* z balíčku *Statistics and Machine Learning Toolbox*. Tato funkce využívá shlukové analýzy, která vstupní data rozdělí do n skupin na základě vzdálenosti v našem případě není vzdálenost pouze trojrozměrná, nýbrž má 9 rozměrů (pozice, barva, normálový vektor).

Funkce má kromě samotných dat a počtu skupin také další vstupní parametry. Mezi ty nejvýznamnější patří počet iterací algoritmu (každá iterace o něco více konverguje k ideálnímu výsledku) nebo způsob výpočtu vzdálenosti, která je pro účely této práce počítána jako euklidovská.



Obr. 17: Segmentace obličejových dat pomocí algoritmu *kmeans*

Na obr. 18 jsou znázorněny jednotlivé výstupní skupiny tak, jak byly rozřazeny algoritmem *kmeans*. Lze si povšimnout, že většina bodů je rozřazena správně. Problém však nastává v okolí očí a úst, kde byla data zařazena nesprávně. Jelikož je přesně známo, o jaká data se jedná a že v určitém okolí nosu bude vždy obličej, je možno tuto chybu manuálně opravit a body, které jsou v předem definovaném okruhu nosu, manuálně přesunout. Výsledek této operace v porovnání se vstupními daty je vidět na obr. 19



Obr. 18: Vizualizace obličejových dat před segmentací (vlevo) a po segmentaci (vpravo)

6.2 Detekce nosu a normalizace dat

V následující části algoritmu je detekována špička nosu. Pokud je možno spolehnout se na předchozí filtraci a segmentaci, je špička nosu v tuto chvíli nejbližším bodem ke kameře (bodem s nejmenší složkou vzdálenosti v ose Z). Takovýto způsob detekce špičky nosu bohužel omezuje maximální úhel, o který může být osoba odkloněna od senzoru. Je to ale známé omezení, je ho tedy možné zachovat vzhledem ke způsobu použití systému rozpoznávání obličeje.

Po samotné detekci jsou obličejové body normalizovány tak, že špička nosu odpovídá středu souřadnicového systému, aby byla následná práce s obličejovými daty jednodušší.

6.3 Extrakce obličejových křivek

Tvar obličeje je velice deskriptivní informací. Víme, že pokud se se zavřenými očima dotkneme obličeje známé osoby, je velká šance, že ji budeme schopni identifikovat. Z hlediska strojového rozpoznání obličeje na základě takovéto informace je nutno obličej rozdělit na rigidní oblast (ta která se s různými výrazy v obličeji příliš nemění) a plastickou (ta která je výrazem v obličeji deformována). Pokud budeme porovnávat pouze ty křivky, které popisují rigidní oblast obličeje, získáme mnohem lepší nezávislost na výrazech.

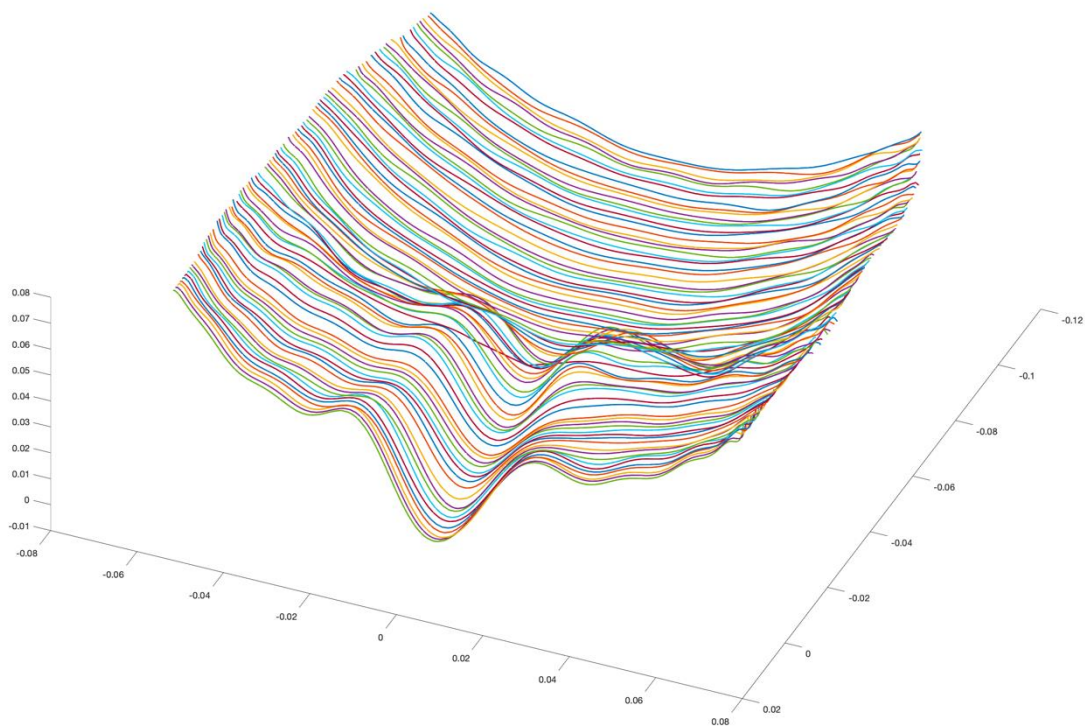
Algoritmus tedy získá vodorovné křivky, které popisují obličej od špičky nosu přes oči až po čelo. Toho je docíleno tak, že body ve stejné rovině proložíme polynomiální rovnicí dostatečného řádu. Řád rovnice je vypočten vztažením k celkovému počtu bodů v dané rovině. Příliš nízký řád rovnice způsobí nedostatečně věrný popis povrchu obličeje, příliš vysoký řád naopak povrch zdánlivě zvlní. Experimentálně byla zvolena závislost popsána rovnicí:

$$n \approx \frac{3 * (k - 400)}{100} + 8$$

n ... řád rovnice zaokrouhlený na celé číslo

k ... počet bodů v rovině

Na obr. 19 je znázorněn výstup extrakce obličejových křivek.



Obr. 19: Extrakce vodorovných obličejových křivek

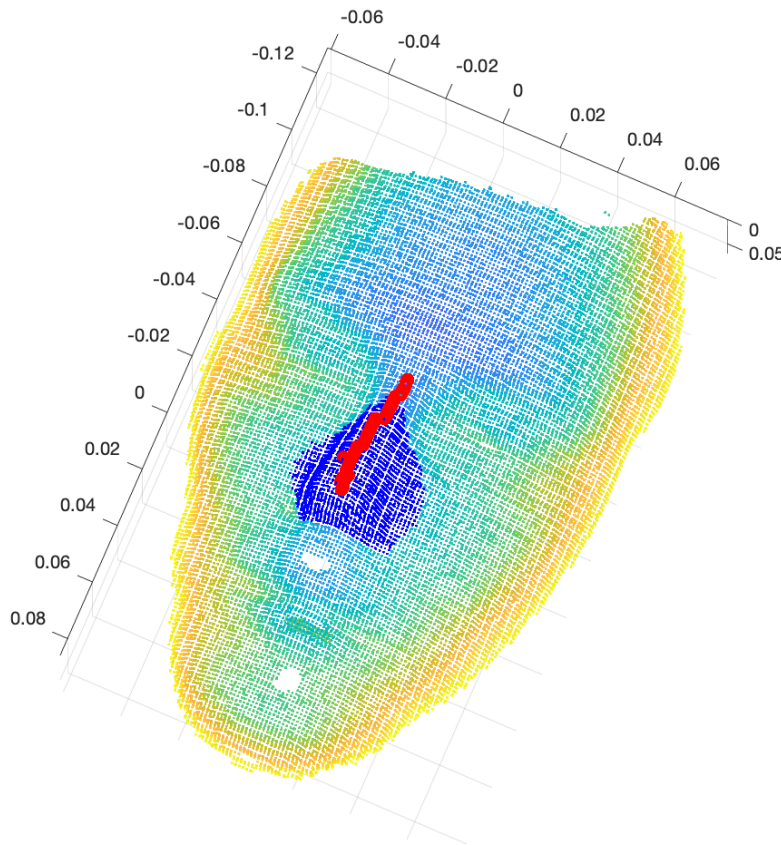
6.4 Extrakce tvaru nosu

Nos je nejrigidnější a zároveň nejvíce deskriptivní místo na obličeji. Je proto velice výhodné zaznamenat jeho tvar. Součástí této práce jsou dva způsoby popisu nosu.

Prvním z nich je extrakce tvaru hřebene nosu, která nese informaci pouze o zakřivení a délce. Tato křivka může být v konečné fázi srovnávána s ostatními obličeji a poskytne cenné informace o rozdílech ve tvaru nosu, což může být rozhodujícím faktorem při rozhodování se mezi dvěma osobami.

Druhým způsobem je extrakce všech bodů, které jsou v definovaném okolí špičky nosu, toto mračno bodů je pak možno zarovnat s body náležícími jinému obličeji a vypočíst střední kvadratickou chybu jejich odchylky.

Na obr. 20 je červenou barvou znázorněna křivka popisující tvar hřebenu nosu, modrou barvou pak body náležící do okolí špičky nosu.



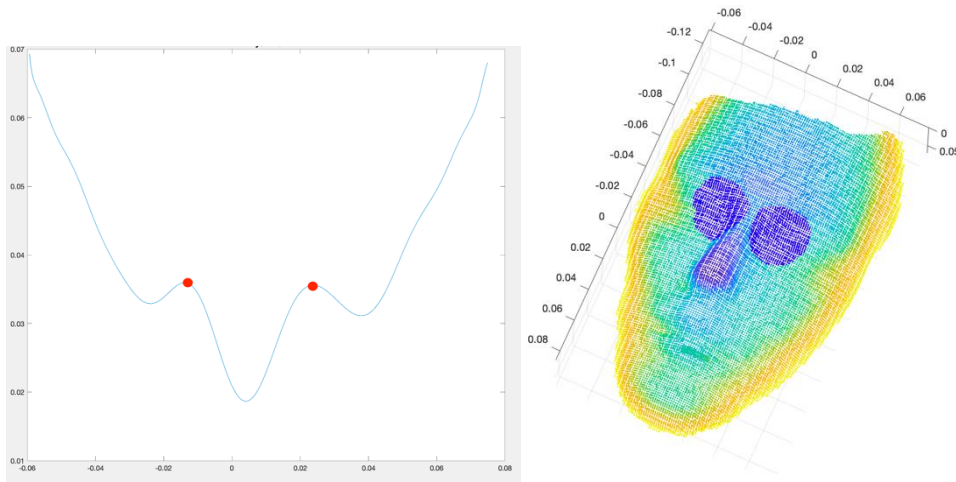
Obr. 19: Extrakce hřebene nosu (červená) a okolí špičky nosu (modrá)

6.5 Extrakce tvaru očních koutků

Oční koutky a jejich okolí jsou dalším velice dobrým bodem pro porovnávání obličejových dat. Jsou z velké části rigidní a jejich detekce není příliš složitá. Pro účely této práce jsou detekovány pouze vnitřní koutky očí.

Detekce probíhá tak, že je vzata rovina ve výšce kořene nosu a její derivací jsou nalezeny body znázorněné na obr. 21 vlevo. Okolí těchto bodů je následně vyfiltrováno podobně jako v kapitole 6.4.

Na obr. 21 vpravo je modrou barvou znázorněn výsledek extrakce.



Obr. 20: Detekce očních koutků z derivace obličejové křivky (vlevo) a jejich okolí (vpravo)

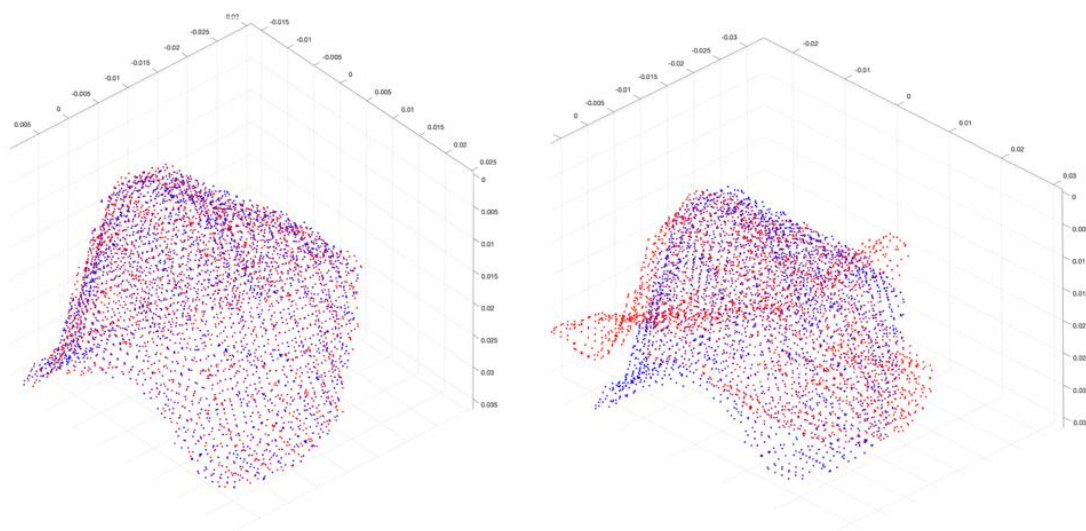
7 POROVNÁNÍ POPISNÝCH BODŮ

Po extrakci popisných bodů pracuje vytvořený skript ve dvou režimech. Prvním z nich je režim uložení dat do databáze, kdy jsou extrahované body uloženy do souboru ve formátu *mat* pro pozdější porovnávání.

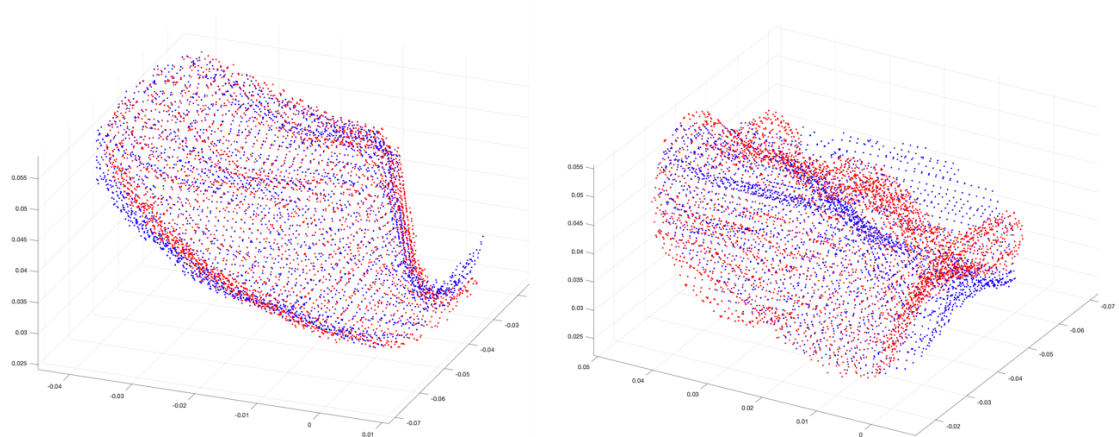
Ve druhém režimu nejsou výsledky nikam ukládány, jsou ale porovnány se všemi záznamy v databázi a je vypočtena míra shody. V této kapitole bude blíže popsáno, jakým způsobem je porovnávání dat provedeno a jakých dosahuje navržený algoritmus výsledků. Kapitola se také blíže věnuje tomu, jak je ovlivněn výsledek v různých nepříznivých situacích.

7.1 Porovnání tvaru nosu a očních koutků

Porovnávání tvaru okolí deskriptivních bodů je zajištěno funkcí *pcregisterndt* z balíčku *Computer Vision Toolbox*. V první fázi jsou jednotlivá mračna bodů zarovnávána pomocí algoritmu NDT (Normal Distribution Transfrom). Jedná se o iterační algoritmus, který minimalizuje celkovou střední kvadratickou odchylku. Po nalezení minima je odečtena finální odchylka, která vypovídá o míře shody. Na obr. 21 je zobrazeno zarovnání okolí nosu patřícího stejné osobě (vlevo) a patřícího různým osobám (vpravo). Porovnání tvaru očních koutků probíhá totožným způsobem a je vyobrazeno na obr. 22.



Obr. 21: Vizualizace porovnání okolí nosu pro výpočet míry shody



Obr. 22: Vizualizace porovnání tvaru očních koutků

8 FÁZE TESTOVÁNÍ

Pro účely otestování navržených algoritmů bylo skenováno celkem 5 obličejů. Základní datová sada každého obličeje obsahuje:

- 2x neutrální výraz
- 1x otevřená ústa
- 1x zamračení
- 1x pohled doleva (přibližně 30° od senzoru)
- 1x pohled doprava (přibližně 30° od senzoru)
- 1x neutrální výraz při výrazně nerovnoměrném osvětlení

Mimo zmíněnou základní sadu pro testování bylo pořízeno mnoho skenů již ve fázi vývoje, které obsahují neutrální výrazy, ale i částečně zakryté obličeje a další ztěžující podmínky.

V této kapitole budou popsány podmínky, za kterých bylo skenování provedeno a zároveň budou definovány podmínky nutné pro správnou funkčnost navržených algoritmů.

8.1 Akvizice základní datové sady

Data byla získána v průběhu jednoho týdne, důraz byl kladen na co možná největší zachování podmínek okolního prostředí.

Senzor Microsoft Kinect V2 je umístěn přibližně ve výšce očí na vrchu monitoru, na kterém je zobrazováno okno programu pro akvizici dat. Obličeje tedy nejsou skenovány ani z podhledu, ani z nadhledu, aby nedocházelo k zakrytí části obličeje nosem. Konzistentní světelné podmínky jsou zajištěny osvětlením umístěným za senzorem tak, aby se na obličeji netvořil žádný stín ani světlejší či tmavší místa. Okno v místnosti je zatemněno žaluziemi.

Osoby stojí v přibližné vzdálenosti 75 centimetrů od senzoru a hledí rovně před sebe. Je zajištěno, aby žádná část těla, která je viditelná, nebyla blíže senzoru, než nos. V průběhu samotné akvizice se osoba nehýbe.

8.2 Testování kvality popisných bodů a jejich extrakce

Data exportovaná z akvizičního programu byla importována do skriptu v prostředí MATLAB a první pořízený sken každé osoby byl skrze první režim algoritmu zpracován a popisné obličejové body byly uloženy do databáze.

Následně, když byla každá z osob uložena v databázi, byly zpracovány zbylé skeny a pomocí druhého režimu algoritmu byly extrahované popisné body porovnány s těmi v databázi. Ukázalo se, že algoritmus je schopen extrakce popisných bodů s vysokou přesností a je jen málo závislý na rovnoměrnosti osvětlení. Algoritmus rovněž zvládá extrahovat popisné body i při odklonu osoby doleva či doprava o přibližně 20°. Bohužel byla prokázána vysoká náchylnost na zakrytí částí obličeje či rotaci (uklonění hlavy k rameni).

Co se týče jednotlivých popisných bodů, nejvíce kvalitním se jeví okolí nosu popsané v kapitole 6.4, které je velice konzistentní při různých výrazech obličeje. Dalším kvalitním ukazatelem jsou oční koutky definované v kapitole 6.5, které jsou ale mírně plastické při změnách ve výrazu obličeje, zejména při mračení.

Naopak méně kvalitním ukazatelem se bohužel prokázala křivka tvaru hřebene nosu, která obsahuje velice málo bodů v porovnání s předchozími ukazateli. Nejméně kvalitním ukazatelem se jeví obličejové křivky z kapitoly 6.3, které jsou nejvíce náchylné na přesnost segmentace, protože někdy obsahují zbytky vlasů.

8.3 Stanovení vztahu pro výslednou procentuální shodu

Aby byl výstup algoritmu co nejvíce uživatelsky přívětivý, byl empiricky stanoven vztah, který s různými vahami agreguje předem zmiňované porovnávání popisných bodů a jeho výstupem je přibližná procentuální shoda.

Při stanovování vztahu byl brán ohled na kvalitu popisných bodů diskutovanou v kapitole 8.2. Z důvodu, že ani dva skeny stejné osoby nejsou vyhodnoceny jako absolutně shodné, byl výsledný vztah o tuto odchylku posunut. Výsledkem porovnávání jsou tedy seřazené osoby z databáze dle procentuální míry shody. Následující rovnice ukazuje, jakým způsobem byly nastaveny váhy pro jednotlivé popisné body.

$$RR = 10 * (800 * noseRmse + 200 * (lEyeRmse + rEyeRmse) - 1)$$

RR ... Míra shody [%]

noseRmse ... střední kvadratická chyba při porovnání okolí nosu

lEyeRmse, rEyeRmse ... střední kvadratická chyba při porovnání očního koutku

8.4 Nutné podmínky pro správnou funkčnost algoritmu

Při vyhodnocování základních testovacích sad a všech ostatních dat, nasbíraných během vývoje, byly vypořádovány různé podmínky, které jsou buď nutné pro samotnou funkčnost algoritmu, nebo pouze eliminují rizika nepřesné extrakce popisných bodů a zvyšují tak jeho úspěšnost.

První podmínkou pro správnou funkčnost je přímý pohled uživatele do senzoru. Z naměřených dat bylo prokázáno, že jakýkoliv odklon větší, než přibližně 20° může velice často způsobit chybějící data za odkloněnou částí nosu, nebo nesprávnou detekci očního koutku odkloněné části obličeje.

Druhou podmínkou je neexistence jakýchkoliv obstrukcí obličeje, jako je například rouška, brýle, či čepice. Ač některé z nich nemusí nutně znamenat, že algoritmus nebude fungovat vůbec, velice často tyto elementy způsobí nesprávnou extrakci popisných bodů a tím velice ovlivní výsledek porovnávání.

Další podmínkou je neutrální výraz v obličeji. Základní testovací sadou byly ozkoušeny výrazy jako mračení, či plně otevřená ústa. Z výsledných dat vyšlo najevo, že některé výrazy v obličeji mají větší vliv na finální výsledek než jiné, někdy až znemožní správnou identifikaci osoby. Pokud jsme schopni zajistit přímý pohled uživatele do senzoru, budeme bezpochyby schopni zajistit i zachování neutrálního výrazu v obličeji.

Posledním aspektem, který byl testován je rovnoměrnost osvětlení obličeje. Ač se zprvu jeví, že nerovnoměrně osvětlené obličeje budou velkým problémem, při finálním testování se neprojeví žádná měřitelná závislost. Bohužel však algoritmus nepracuje správně při velice nízkých intenzitách osvětlení, kde je hlavním problémem segmentace obličeje.

9 POUŽITELNOST SYSTÉMU V PRAXI

Navržený systém je schopný, jak bylo dokázáno v kapitole 8, spolehlivě rozpoznávat obličeje, pokud jsou dodrženy podmínky specifikované v kapitole 8.4. Bohužel, vzhledem k nouzovému stavu, který probíhal v době vypracovávání této bakalářské práce, nebylo možné získat velké množství obličejových dat pro testování. Z tohoto důvodu tato práce bohužel neposkytuje přesně vypočtené chyby pro porovnání s ostatními systémy rozpoznávání obličeje.

Z omezeného testování je usuzováno, že v současné podobě je systém schopen rozpoznávat osoby v malých skupinách (nižší desítky lidí), kde je například použitelný pro kontrolu docházky na vrátnici firmy, nelze ho však použít ke kontrolovanému vstupu do prostor s omezeným přístupem. K takovému účelu by musel být nadále zdokonalován, testován a certifikován.

Běžná výsledná shoda pro stejné osoby bývá okolo 97 %, druhá nejlepší shoda (s nesprávnou osobou) se pohybuje pod 85 %. Pokud jsou dodrženy všechny podmínky, má systém úspěšnost přibližně 96 % pro správnou identifikaci osoby, je však stále potřeba dozoru pracovníkem, který bude jeho výstup kontrolovat a asistovat v případě problémů.

Co se týče rychlosti celého procesu skenování, extrakce popisných bodů a rozpoznání obličeje, systém dosahuje doby skenování přibližně 3 sekundy. Po exportu a vložení dat do algoritmu v prostředí MATLAB zabere extrakce popisných bodů přibližně 25 sekund, přičemž 80 % z tohoto času je stráveno na úvodní filtraci dat. Doba porovnávání dat s ostatními uživateli je závislá na jejich počtu a je pro každé porovnání přibližně konstantní (500 milisekund).

ZÁVĚR

Bakalářská práce se věnovala problematice strojového rozpoznávání obličejů za pomoci levného senzoru Microsoft Kinect V2. Byla provedena rešerše stávajících řešení s přihlédnutím k výhodám i nevýhodám každého z nich.

Metody popisu 3D obličejových dat byly rozděleny do 3 skupin, podle způsobu náhledu na obličejová data. Metody holistického přístupu nahlíží na vstupní data jako na celek a provádí na základě nich úsporný popis obličeje. Naopak metody lokálního přístupu vyhledávají v obličejových datech pouze charakteristické body, křivky či plochy. V praxi se často využívají různé kombinace metod holistického i lokálního přístupu. Zvyšuje se takto bezpečnost a přesnost celého systému, či odolnost oproti různým rušivým vlivům.

Kapitola 4 věnovala pozornost senzoru Microsoft Kinect V2 a byly v ní představeny jeho funkce a hardwarové limity. Součástí této kapitoly bylo i nahlédnutí do sady vývojových nástrojů, zejména pak do knihovny Kinect Fusion, která je základním stavebním kamenem způsobu, jakým je v této práci prováděna 3D rekonstrukce obličeje a je tedy třeba ji porozumět.

V kapitole 5 byl popsán obslužný program pro získání 3D obličejových rekonstrukcí uživatele a jejich následný export ve formátech PLY a CSV pro pozdější extrakci popisných bodů a porovnání s ostatními uživateli.

Proces extrakce popisných bodů byl popsán v kapitole 6. Tato kapitola také obsahovala množství ilustračních obrázků znázorňujících vzhled extrahovaných bodů. Kapitola 7 obsahovala popis způsobu, jakým byla extrahovaná data porovnávána mezi sebou. Kapitola 8 popisovala fázi testování, při které byly nasbírány základní datové sady a systém na nich byl otestován. Během tohoto procesu byly přiřazeny váhy pro jednotlivé popisné body, dle jejich robustnosti. Od použití nějakých popisných bodů, získaných v předchozích částech algoritmů, bylo dokonce upuštěno. Výsledkem této kapitoly byla sada podmínek a doporučení pro správnou funkčnost navrženého algoritmu.

Kapitola 9 shrnula výsledky dosažené v této práci a navrhla možné použití systému v praxi.

Literatura

- [1] JACKSON, Aaron S., Adrian BULAT, Vasileios ARGYRIOU a Georgios TZIMIROPOULOS. *Large Pose 3D Face Reconstruction from a Single Image via Direct Volumetric CNN Regression* [online]. Venice, Italy, 2017 [cit. 2019-12-31]. Dostupné z: <https://ieeexplore.ieee.org/abstract/document/8237379>. The University of Nottingham, Kingston University, UK.
- [2] SOLTANPOUR, Sima, Boubakeur BOUFAMA a Q.M. Jonathan WU. *A survey of local feature methods for 3D face recognition* [online]. School of Computer Science, University of Windsor, Canada, 2017 [cit. 2019-12-31]. Dostupné z: <https://www.sciencedirect.com/science/article/pii/S0031320317303072>
- [3] AMOR, Boulbaba Ben, Karima OUJI, Mohsen ARDABILIAN a Liming CHEN. *3D Face recognition by ICP-based shape matching* [online]. Lyon Research Center for Images and Intelligent Information Systems, 2005 [cit. 2019-12-31]. Dostupné z: <https://liris.cnrs.fr/Documents/Liris-1963.pdf>
- [4] RUSS, Trina, Chris BOEHNEM a Tanya PETERS. *3D Face Recognition Using 3D Alignment for PCA* [online]. New York, NY, USA, 2006 [cit. 2019-12-31]. Dostupné z: <https://ieeexplore.ieee.org/abstract/document/1640920>. Sandia National Labs Albuquerque, NM; University of Notre Dame Notre Dame, IN.
- [5] GERVEI, Omid, Ahmad AYATOLLAHI a Navid GERVEI. *3D Face Recognition Using Modified PCA Methods* [online]. Iran University of Science and Technology, Tehran, Iran, 2010 [cit. 2019-12-31]. Dostupné z: <https://publications.waset.org/5789/pdf>
- [6] LI, Huibin, Di HUANG, Pierre LEMAIRE, Jean-Marie MORVAN a Liming CHEN. *Expression robust 3D face recognition via mesh-based histograms of multiple order surface differential quantities* [online]. Brussels, Belgium, 2011 [cit. 2019-12-31]. Dostupné z: <https://ieeexplore.ieee.org/abstract/document/6116308>. Université de Lyon, CNRS.
- [7] COLOMBO, Alessandro, Claudio CUSANO a Riamondo SCHETTINI. *3D face detection using curvature analysis* [online]. Università degli Studi di Milano-Bicocca, Milano, Italy, 2005 [cit. 2019-12-31]. Dostupné z: <https://www.sciencedirect.com/science/article/pii/S0031320305003791>. Université de Lyon, CNRS.

- [8] SMEETS, Dirk, Peter CLAES, Jeroen HERMANS, Dirk VANDERMEULEN a Paul SUETENS. *A Comparative Study of 3-D Face Recognition Under Expression Variations* [online]. Katholieke Universiteit Leuven, Belgium, 2012 [cit. 2019-12-31]. Dostupné z: <https://ieeexplore.ieee.org/abstract/document/6101586>
- [9] SAMIR, Chafik, Anuj SRIVASTAVA a Mohamed DAOUDI. *Three-Dimensional Face Recognition Using Shapes of Facial Curves* [online]. Los Vaqueros Cir, Los Alamitos, United States, 2006 [cit. 2019-12-31]. Dostupné z: <https://ieeexplore.ieee.org/abstract/document/1704840>
- [10] AHDID, Rachid, Khaddouj TAIFI, Said SAID, Mohamed FAKIR a Bouzid MANAUT. *Automatic Face Recognition System using Iso-Geodesic Curves in Riemanian Manifold* [online]. Sultan Moulay Slimane University Beni Mellal, Morocco, 2017 [cit. 2019-12-31]. Dostupné z: <https://ieeexplore.ieee.org/abstract/document/8361546>
- [11] LEI, Yinjie, Mohammed BENNAMOUN, Munawar HAYAT a Yulan GUO. *An efficient 3D face recognition approach using local geometrical signatures* [online]. University of Western Australia, Australia, 2014 [cit. 2019-12-31]. Dostupné z: <https://www.sciencedirect.com/science/article/pii/S0031320313003166>
- [12] NEWCOMBE, Richard A., Shahram IZADI, Otmar HILLIGES, et al. *KinectFusion: Real-Time Dense Surface Mapping and Tracking* [online]. Basel, Switzerland, 2011 [cit. 2020-01-01]. Dostupné z: <https://ieeexplore.ieee.org/document/6162880>. Microsoft Research.
- [13] SMEETS, Dirk, Johannes KEUSTERMANS, Dirk VANDERMEULEN a Paul SUETENS. *MeshSIFT: Local surface features for 3D face recognition under expression variations and partial data* [online]. Katholieke Universiteit Leuven, Belgium, 2012 [cit. 2020-01-01]. Dostupné z: <https://www.sciencedirect.com/science/article/pii/S1077314212001324>
- [14] HAOYU, Chen. *MeshSIFT: Local surface features for 3D face recognition under expression variations and partial data* [online]. University of Oulu, Finland, 2017 [cit. 2020-01-01]. Dostupné z: <http://jultika.oulu.fi/files/nbnfioulu-201706022490.pdf>

Seznam příloh

Příloha 1. – Elektronicky odevzdáno do IS VUT FEKT.

Tato příloha obsahuje:

- Zdrojový kód obslužného programu senzoru Microsoft Kinect v2
- Skript v programovacím prostředí MATLAB
- Vzorový obličejový sken

Příloha 2. – USB Flash disk, který je součástí tištěné práce

Tato příloha obsahuje:

- Zdrojový kód obslužného programu senzoru Microsoft Kinect v2
- Skript v programovacím prostředí MATLAB
- Vzorový obličejový sken
- Testovací sada obličejových skenů